AD_____

Award Number: W81XWH-09-1-0068

TITLE: Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent

PRINCIPAL INVESTIGATOR: Mark Pegram, M.D

CONTRACTING ORGANIZATION: The Leland Stanford Junior University
Stanford, CA 94305

REPORT DATE: July 2013

TYPE OF REPORT: Final report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2013	Final	1 June 2009-30 April 2013

4. TITLE AND SUBTITLE
Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-09-1-0068

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Mark Pegram, M.D.

E-Mail: mpegram@stanford.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S)

The Leland Stanford Junior University
Stanford, CA 94305

Stanford University
265 Campus Drive Rm G2005
Stanford, CA, 94305

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent

14. ABSTRACT

Background: Breast cancer (BC) is the second leading cause of cancer death among African-American (AA) women, with mortality 20% greater than that in Caucasians (Cauc). However, the basis for such disparity remains an enigma. Recent observations from our laboratory suggest the involvement of unidentified genes contributing to AA BC risk. Matched tumor and normal FFPE samples from Cauc and AA patients were obtained from the UM /Sylvester Breast Tissue Bank (UM/S BTB) under an IRB-approved protocol. Based on analysis of 22,000 transcripts, ethnic specific gene expression patterns were identified that may provide important new insights into molecular mechanisms of ethnic subtype differences in clinical outcomes. We propose to extend these preliminary findings to a large African tumor bank [available via collaboration between Drs. Peter A. Bird (Kijabe, Kenya) and Mark Pegram (UM Sylvester).] Additionally, we propose to analyze chromosomal alterations associated with gene expression differences utilizing array cGH (in collaboration with Alan Ashworth, England). This work will contribute to development of rationale designs of preventive, predictive and therapeutic measures for BC in different ethnicities, and thus, a significant reduction in current ethnic-specific disparities in BC incidence, morbidity and mortality.

Hypothesis: ***Discrete genomic alterations and gene expression changes will be identified and shared between triple negative tumor specimens within an ethnic group, i.e., North Americans/African decent and Kenya.*** **Aim I:** ***Analyze and compare genome-wide differences in gene expression in BC samples of AA ancestry vs. native African (Kijabe) samples (Drs. Pegram, Baumbach, Bird, Halsey). Aim II: Investigate possible chromosomal alterations associated with gene expression differences*** (***Drs. Pegram, Baumbach, Ashworth).*** **Aim III:** ***Analyze ancestry of each sample using a panel of ancestry-informative DNA markers (Drs. Kittles, Baumbach).***

Synergy Statement: The proposed investigations are highly synergistic. This study will also allow for the first direct comparison of gene expression/genomic copy number data in triple negative tumor specimens across Americans of African decent and Kenyan East Africans. We will correlate all experimental data with a spectrum of clinical data available on study subjects, and apply covariate modeling and logistic regression analysis to determine possible correlations between genomic signatures, genomic changes, clinical tumor characteristics and outcomes/ response measures among and across ethnic groups.

15. SUBJECT TERMS

Triple negative breast cancer, Ethnic disparities, Breast cancer amongst African-Americans and Africans, Gene expression profiling, Array CGH

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Mark Pegram
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		19	**19b. TELEPHONE NUMBER** *(include area code)* *305-4484-8949*

Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent

Table of Contents

Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent

1) INTRODUCTION

The advent of microarray technology has enabled the robust, high throughput analysis of disease specific transcriptomes, including those in breast tumor specimens. Indeed, the molecular classification of breast cancer has been revolutionized by the advent of gene expression profiling. However, currently available commercial microarray design focuses on the most commonly known and characterized genes from all body tissues, therefore only a subset of genes on a generic microarray will yield informative results for any tissue-specific study. Moreover, since the transcriptome of a given tissue contains tissue/disease-specific splice variation as well as non-coding RNAs, many important transcripts solely expressed in the tissue of interest will not be represented. One innovative solution to this problem that we will utilize in this project is to exploit custom breast cancer-specific arrays developed by our collaborators at Almac Diagnostics. With tens of thousands of transcripts not found on generic arrays, specificity of differential gene expression patterns will be significantly enhanced. Furthermore, the use of expression array technology historically has been dependent upon the availability of intact RNA from fresh frozen tumor tissue for analysis, thus study of the many large retrospective cohorts with annotated clinical follow-up has not been possible. RNA extracted from FFPE samples tends to have shorter median length from 3' to 5' and the detection of these transcripts on generic array platforms is rarely successful. However, using an innovative approach we have recently successfully tested novel array probes specifically designed to detect partially degraded RNA from formalin-fixed, paraffin-embedded (FFPE) breast tumor material from samples at the University of Miami. The use of a probeset with extreme 3' sequence mitigates this previous technical limitation, and thus is considered highly innovative.

Another innovation in this study is the genomic analysis of a published East African breast cancer cohort, the largest of its kind from the region. Importantly, the integration of high density array cGH technology with the expression array data is highly innovative (to our knowledge, the first study of this kind in a native African, or even African American cohort). This approach will allow identification of ethnic specific copy number variation and loss of heterozygosity, and their relation to gene expression changes. Finally, the incorporation of an ancestry marker panel makes this a particularly novel study which is sure to produce data of interest to the community. Our eventual goal will be to develop further understanding of biology of disease, prognostic biomarkers, and eventually, the targets for therapeutics for ethnic-specific subgroups in breast cancer.

2) BODY

Characteristics of Study Population

Breast cancer is the second leading cause of cancer death among African-American (AA) women (1). Mortality is 20% greater than that in Caucasian (Cauc) women, and is partially attributed to more aggressive disease and poorer prognosis. In addition, AA women $\leq$50 years have the highest rate of new breast cancer cases in the US (1,2). General consensus exists that AA women of all ages are more likely to have poorly differentiated breast cancer, which is likely to occur at an earlier age, be ER and PR negative, and to have a higher proliferative fraction - all factors associated with more aggressive tumors (2). Therefore, the prognosis in AA patients is worse, even adjusted for stage of presentation. Ethnic-specific differences in response to adjuvant therapy have also been reported (3,4). Taken together, the cumulative data suggests that intrinsic, ethnic-specific, and biological/genetic differences contribute to disparities in breast cancer morbidity and mortality**.**

A recent study by Bird et al (5) focused on a cohort of BC patients from the Kijabe Hospital in Kenya and reported a very low frequency of hormone receptor expression: 24% ER-positive and 34% ER-or PR-positive tumors. Compared to breast cancer in Western or Cauc populations, the Kijabe patients have a high proportion of poorly differentiated, advanced cancers and irrespective of disease stage, were much less likely to be hormone sensitive (ER and PR negative). Overall, the possibility of inherently more aggressive tumor biology, coupled with low hormone receptor sensitivity, may represent manifestations of modified biology in African populations. This study further characterizes the tumors in a Kijabe clinical cohort. A set of 55 residual pathology tissue blocks were obtained from Dr. Peter Bird in Kenya. These samples were selected by Dr. Bird on the basis on of being ER or PR negative by clinical testing. Her2 testing had not previously been done on any of the samples. For all 55 cases sections were cut and stained by immunohistochemistry for ER, PR and Her2. The stained slides were evaluated by UM pathology to find all cases which were triple negative (negative for ER, PR and Her2). The table below shows the results of immunohistochemical staining.

Table 1. Hormone Receptor Status of Kijabe Breast Cancer Cohort

Receptor Status	No. of Samples
ER-/ PR-/ Her2-	31
ER-/ PR-/ Her2+	10
ER-/ PR+/ Her2-	0
ER+/ PR-/ Her2-	8
ER+/ PR-/ Her2+	2
ER+/ PR+/ Her2-	2
ER+/ PR+/ Her2+	2

Staining showed that 31 out of the 55 samples (or 56%) were triple negative breast cancer samples. These 31 samples were selected for use in this project. Of the remaining cases 12 were positive for Her2 staining with nine samples (or 16%) with strongly positive score of +3. Her2 staining is interpreted on the maximum area of staining intensity as follows: 0 = no staining; +1 = weak, incomplete membranous staining; +2 = moderate, complete membranous staining of at least 10% of invasive tumor cells; and +3 = strong membranous staining of at least 10% of invasive tumor cells. Cases interpreted as 0 or +1 are considered negative, and cases interpreted as +2 or +3 are considered positive (Figure 1). Many of the cases with Her2 +3 staining appeared to be advance stage aggressive cancers. In general the cohort of 55 samples reflects the overall advanced stage and high incidence of hormone receptor negative cases seen in both African and African-American breast cancer cases.

Gene Expression Array Studies

The 31 triple negative Kijabe samples were then matched with an equal number of African-American samples from South Florida. Samples were obtained from residual pathology tissue blocks for cases confirmed to be triple negative by immunohistochemistry. Once samples are identified RNA is extracted and expression profiling done using the Almac Diagnositics Breast Cancer Disease Specific Array (DSA). Quality control checks are completed at each step of the process, for RNA quality spectrophotometer and the Agilent Bioanalyzer are used, the Bioanalyzer provides more sensitive qualitative analysis from less RNA than other traditional methods. The bioanalyzer uses a fluorescent assay and electrophoretic

separation to evaluate RNA samples qualitatively. The software creates a graph called an electropherogram, high quality RNA electropherograms exhibit two primary characteristics. First, clear 28S and 18S peaks and secondly, there should be low noise between the peaks and minimal low molecular weight contamination. Samples meeting these criteria are then processed for hybridization to the DSA array.

Data resulting from the DSA hybridizations are checked for quality control by first looking at the distribution of the sample data (histogram of normalized intensity values) will be assessed to determine what statistical tests will be applied in later stages of analysis. The data had a normal distribution so K-Mean Clustering was performed: In K-Mean clustering groups are created which shows the relationships among the expression levels of conditions or samples. This allows identification any spurious samples, a particular concern when replicates are included in an experiment. K-Mean is used because of prior knowledge of samples condition as being either from tumor or normal tissue.

In addition to K-Mean Clustering Principal Components Analysis is also performed: This is a decomposition technique that produces a set of expression patterns known as principal components. Linear combinations of these patterns can be assembled to represent the behavior of all of the genes in a given data set. Although not a clustering technique the aim of PCA is similar to that of clustering. It is a tool to characterize the most abundant themes or building blocks that reoccur in many genes in the experiment.

Two-dimensional hierarchical clustering analysis was performed to examine the gene expression patterns across samples groups at intensity level. The result was shown in a heatmap (see example in Figure 1 below). From the heatmap, data from the Kenyan (Native African) tumor samples are clearly different in expression pattern from normal tissue from African-Americans.

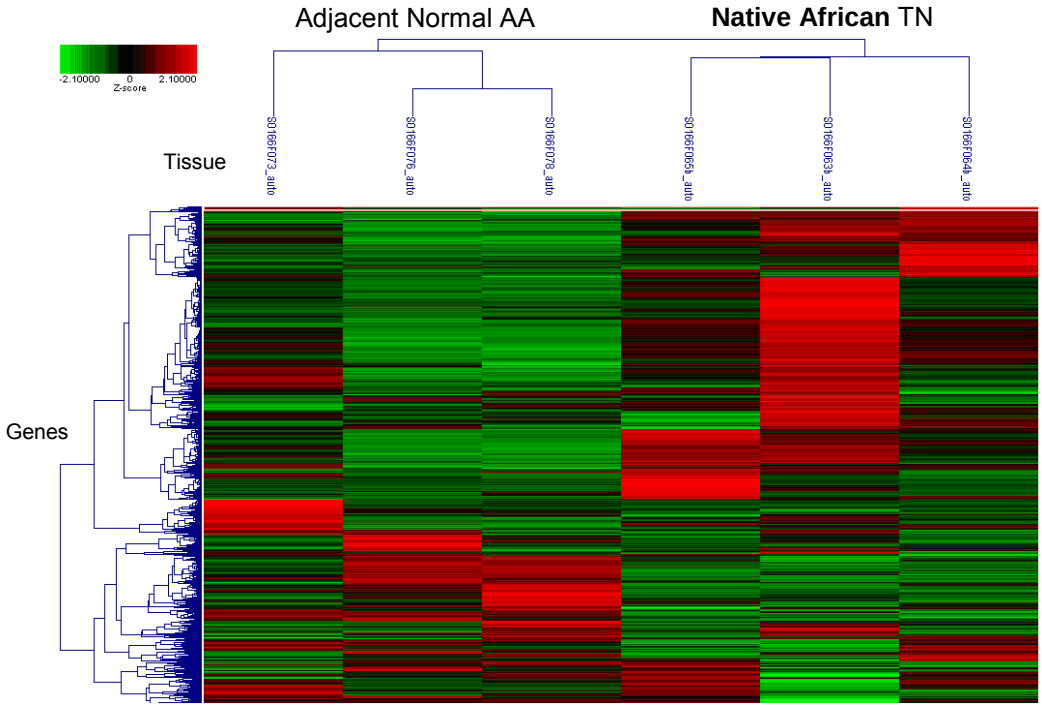


Figure 1. Gene Expression Pattern of Native African Triple Negative Breast Cancer and African-American Adjacent Normal Breast Tissue

Also, following quality control measures, Stringent and Less Stringent Gene Lists are generated from the expression data. For the differentially expressed genes, genes with intensity greater than the background intensity plus the 3 standard deviations are retained as presence call in the data. For stringent genes, genes with intensity greater than 2X background intensity were retained in the sample group. For the Differentially Expressed Gene List, cut-offs of a p-value of 0.01 in 2way ANOVA and paired-t-tests are applied. Sequences with significant statistical confidence (p-value < 0.01 in both tests) were retained in these differentially expressed gene lists. These genes/transcripts are subjected to pathway analysis in Metacore GeneGo program.

New Gene Expression Array Data: Kenyan Cohort

The triple negative Kenyan samples identified above in Table 1 were cut and sent to Almac Diagnostics for RNA extraction. Twenty-four of these samples yielded RNA of sufficient quantity and quality to be hybridized to the Breast Cancer DSA arrays. Quality Control analysis of the arrays was completed and 15 of the arrays showed a greater than 20 percent presence call rate, these samples were used in further data analysis. The Kenyan samples were analyzed in comparison to seven African-American samples that were processed in parallel with the African samples and run on the Breast Cancer DSA. The data from the arrays were RMA normalized and Log2 transformed. A QA/QC and PCA plot (Figure 2 below) shows significant differences between Kenyan and African American samples. Since array experiments were performed in one batch at same time, this variation may represent genuine variation between the populations.

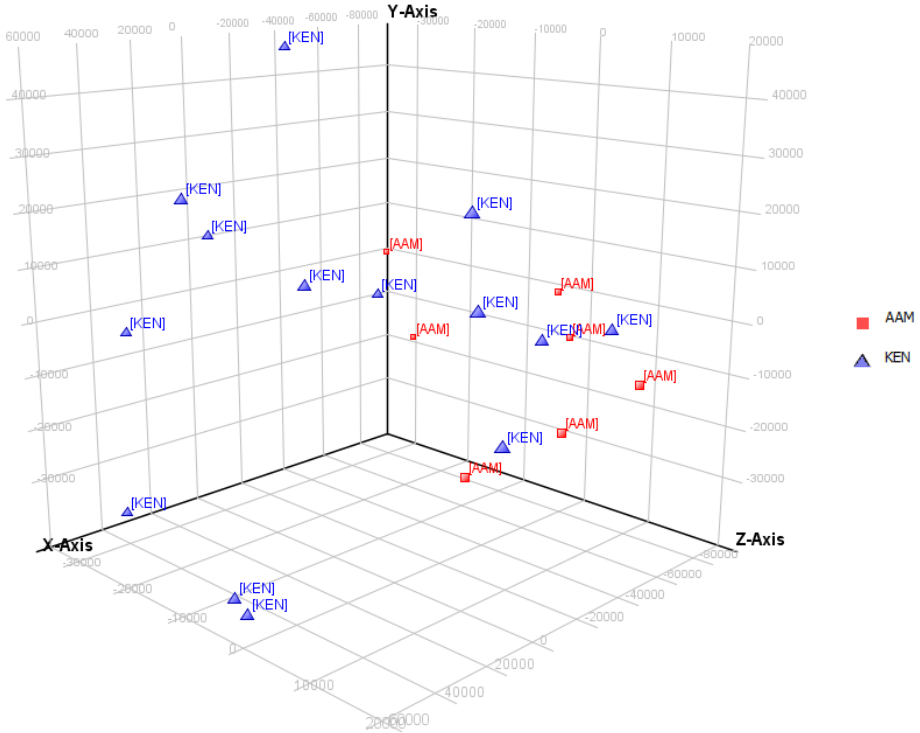
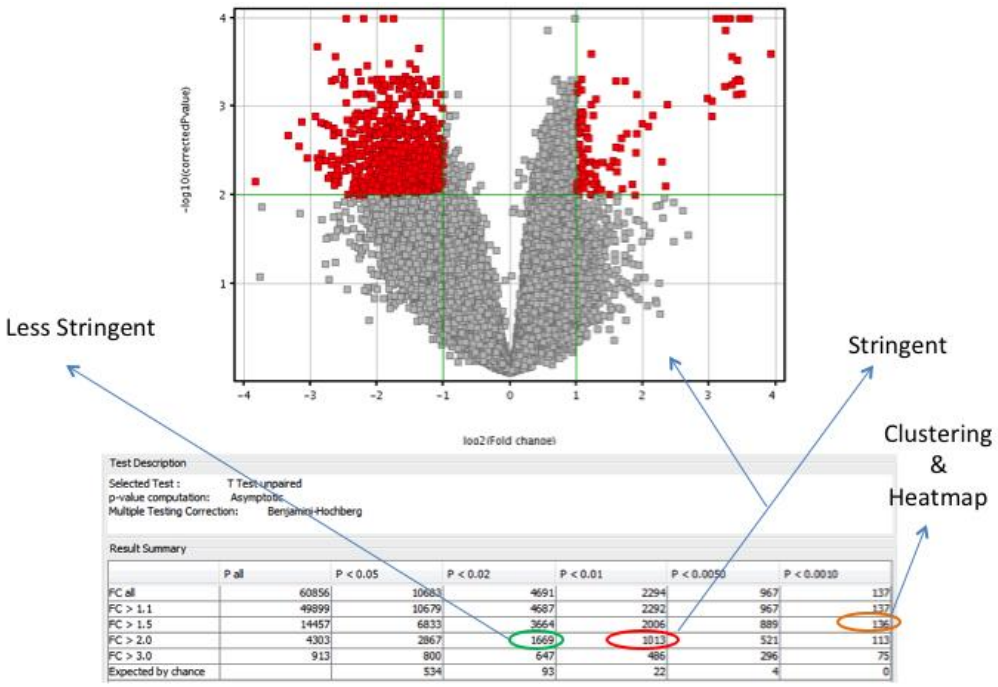


Figure 2. 3D PCA plot of Kenyan & African American Samples. Kenyan samples are indicated by the blue triangles and African-American samples by the red squares.

Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent

An unpaired student's T-test was performed between Kenyan and African-American samples and p-values were corrected for Multiple Testing using Benjamini-Hochberg method. Three sets of differentially expressed genes/probes were extracted using different thresholds. A Stringent group with a p-value ≤ 0.01 & fold change ≥ 2.0 which includes 1013 probes, a Less Stringent group with a p-value ≤ 0.02 & fold change ≥ 2.0 which includes 1669 probes and finally a separate list of 136 differentially expressed probes that was used to generate cluster and heatmap of the most significant genes: p-value ≤ 0.001 & fold change ≥ 1.5. The volcano plot from the unpaired T-test is shown below in Figure 3 and the heatmap resulting from cluster analysis in Figure 4.



	P all	P < 0.05	P < 0.02	P < 0.01	P < 0.0050	P < 0.0010
FC all	60856	10683	4691	2294	967	137
FC > 1.1	49899	10679	4687	2292	967	137
FC > 1.5	14457	6833	3664	2006	889	136
FC > 2.0	4303	2867	1669	1013	521	113
FC > 3.0	913	800	647	486	296	75
Expected by chance		534	93	22	4	0

Figure 3. Volcano Plot from Unpaired T-test btween the Kenyan and African-American samples. The table shows the p-values and fold changes which result in the three sets of differentially expressed genes.

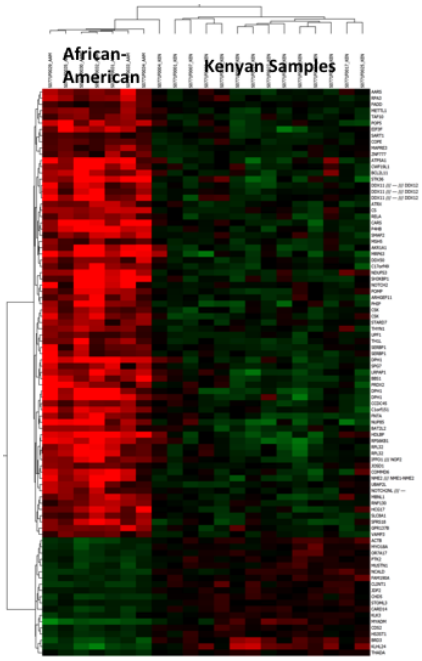


Figure 4. Heatmap of the Kenyan and African-American samples. The attributes of the heatmap are: Clustering Algorithm: Hierarchical; Clustered On: Entities and Conditions; Similarity Measure: Pearson Centered.

Preliminary analysis of the differentially expressed gene sets has been completed using and enrichment analysis workflow in MetaCore portion of GeneGo software. Metacore an integrated knowledge database and software suite for pathway analysis of experimental data and gene lists. To date the interpretation of the pathway analysis is not complete but some of the interesting pathways which appear to be differentially expressed between the Kenyan an African-American samples include: the AKt signaling pathway, the transport of RAN regulation pathway and Transcription Receptor mediated HIF regulation.

Further analysis of the Kenyan data is ongoing including a differential gene expression comparison of the Kenyan samples to a larger cohort of African-American triple negative DSA data.

Copy Number Variation (aCGH) arrays

As cancer cells develop, they undergo dramatic DNA rearrangements such as chromosome loss (Loss of Heterozygosity; LOH) or duplication or translocation. We are using high density CGH arrays to analyze genome wide variation to assess whether gene expression differences may be due to chromosomal alterations.

aCGH is performed using the Breakthrough Breast Cancer 32K tiling path microarray platform, which has a complete coverage of the whole genome with a resolution of 50kb. Details of labelling, hybridization, washes, image acquisition, data pre-processing, normalization and analysis were previously reported (6). Data analysis of these arrays is as follows, cases with >10% of clones missing and clones for which data are not available in ≥10% of cases will be excluded. Log2 ratios will be normalized for spatial and intensity dependent biases using a two-dimensional loess regression followed by a BAC-dependent bias correction (6). The final dataset of BAC clones with unambiguous mapping information according to

the build hg19 of the human genome is used for further analysis. A categorical analysis is applied to the BACs after classifying them as representing gain, loss, or no-change according to their smoothed Log2 ratio values. Threshold values have already been defined in previous studies (6). These thresholds accurately identify low level gains, which are defined as a smoothed Log2 ratio of between 0.12 and 0.45, corresponding to approximately 3-4 copies of the locus, whilst gene amplifications are defined as having a Log2 ratio > 0.45, corresponding to more than 5 copies (see Figure 5 below for example aCGH results).

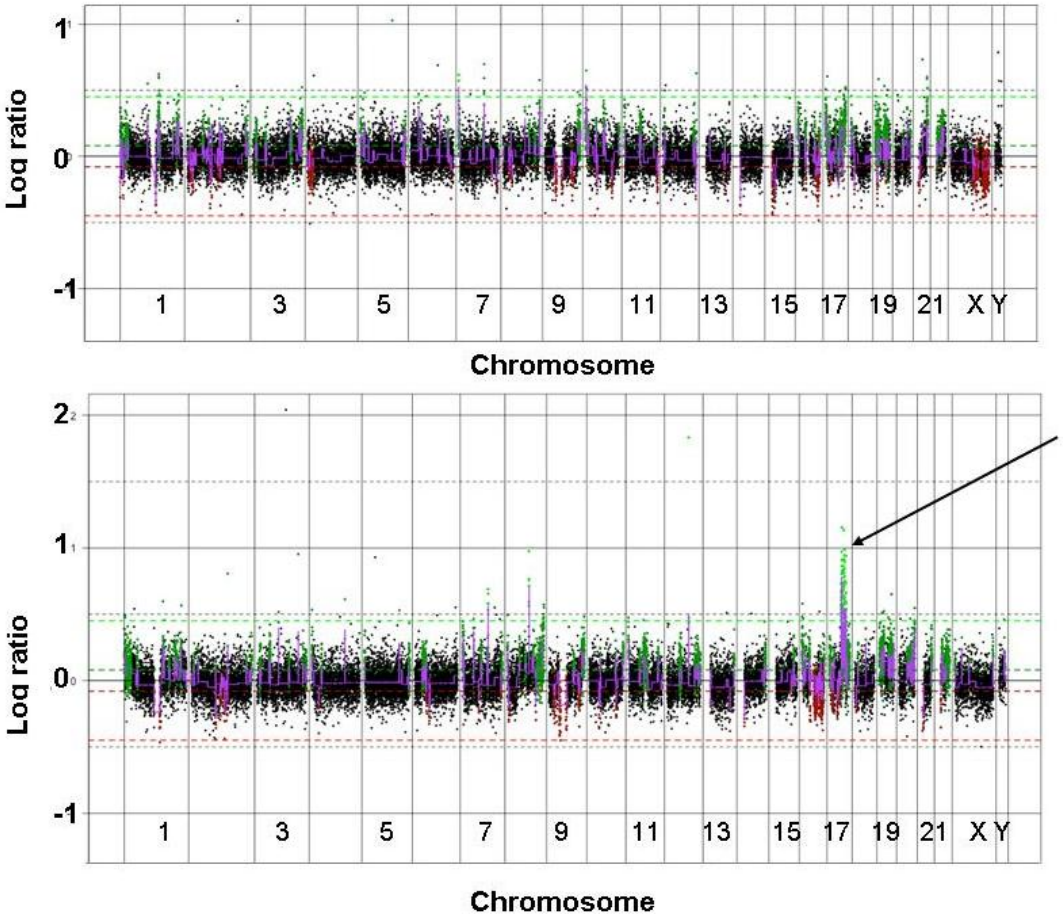


Figure 5. Copy number changes in two Kijabee Kenyan samples from project cohort. Genome plots from aCGH results with log2 ratios for each clone (*Y axis*) plotted according to chromosomal location (*X axis*). *Horizontal line,* centromere. *Green,* gains; *red,* losses. *Top Panel:* sample 1886, showing no large-scale changes in DNA copy number. *Bottom Panel:* sample 1887, showing a single large alteration, a duplication in chromosome 17 (see arrow).

Cell Sorting of Kenyan FFPE Samples

Despite the initial promising pilot data shown in Figure 5, we have subsequently faced a significant technical challenge -- as DNA extracted from most of the Kijabe Kenyan cohort tumor samples does not yield material of sufficient quality to execute the proposed aCGH experiments in the Ashworth laboratory. Consequently, as an alternative solution to this pitfall, a new effort was made to obtain DNA from the Kenyan FFPE samples through collaboration with Dr. Mike Barrett of TGen. Dr. Barrett's lab is expert in extraction of DNA from FFPE sample and cell sorting. Five Kenyan FFPE blocks were chosen

which showed a large section of tissue, 50 micron sections were cut from the blocks and transferred to the Barrett lab at TGen. The tissue was separated from the paraffin and subjected to flow sorting to identify cell populations which represented differences in ploidy. All five samples passed initial QC and resulted in either two or four cell populations. The cells sorted into the following cell population types: 2N, 4N-Diploid, Anuploid or 4N-Anuploid as seen in Table 2 below.

Table 2. Cell Sorting of Kenyan FFPE Samples

ID #	Sample Name	Received	FACS/Sorted	2N	4N-DIP.	AN.	4N-AN.
1	08 3958- C	8/1/2012	9/26/2012	440,000	NA	150,000	NA
2	08 4691	8/1/2012	9/26/2012	180,000	NA	150,000	NA
3	09 630	8/1/2012	9/27/2012	330,000	50,000	60,000	50,000
4	09 3188-B	8/1/2012	9/27/2012	230,000	NA	60,000	NA
5	08-45354	8/1/2012	9/27/2012	390,000	50,000	NA	NA

Sample 3 in Table 2 (09-630) was the only sample to show four distinct cell populations as shown in Figure 6 below.

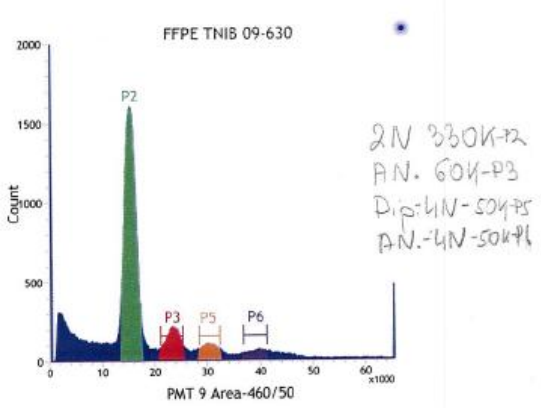


Figure 6. Cell Sorting of Sample 09-630. The first peak, P2 is the 2N population, the second peak AN, the third peak diploid-4N and the fourth peak is anuploid-4N.

After cell sorting each individual cell population was used for a separate DNA extraction. The isolated DNA was hybridized to copy number arrays. However, the hybridizations produced very weak signal. Attempts to repeat the procedure on two of the samples produced the same results. The presumed cause for the weak signal was that the DNA more fragmented than that normally used in the hybridizations as any DNA smaller than 60 bases after labeling will produce a weak signal and limited data.

This experiment was the third different attempt in three different lab settings to obtain usable DNA from the Kenyan FFPE samples. It is possible that the extremely fragmented nature of the DNA is the reason we were unable to obtain good results in copy number array experiments as well as the ancestory informative marker studies (cf. below).

Ancestry Informative Markers (AIMS)

This set of genome-spanning SNPs provides a rich source of information for examining admixture in African Americans these are used rule out spurious results due to underlying population stratification. These portion of the project is to genotype 100 carefully selected ancestry informative markers for all the AA samples. 100 autosomal SNP AIMs are genotyped using the Sequenom MassARRAY platform and iPLEX™chemistry. iPLEX assays were designed utilizing the Sequenom Assay Design software,

allowing for single base extension (SBE) designs used for multiplexing. Individual SNP genotype calls are then generated using Sequenom TYPER software, which automatically calls allele specific peaks according to their expected masses. Quality control checks include genotyping in duplicate multiple samples (10%) in each plate of DNA; cases and unaffected controls are gridded together in each plate to avoid any systematic biases between plates. Individual African ancestry will be estimated from the genotype data using the Bayesian Markov Chain-Monte Carlo (MCMC) method implemented in the program STRUCTURE version 2.1. STRUCTURE will be run under the admixture model using prior population information and independent allele frequencies. Ancestry estimates generated from these AIMs will allow for accurate estimates of European ancestries in our AA subjects, allowing us to utilize individual ancestry estimates as additional covariates in overall experimental analyses. Data from a separate control sample set of 112 African-American samples from South Florida has been completed and shows a range of 62%-98% African ancestry with a mean of approximately 72% African ancestry. This cohort represents a random sample of South Florida African-Americans and should reflect the overall range of African ancestry in the African-American samples used in the gene expression/CNV studies in the current project.

New Research Accomplishments 2013

Mulit-Ethnic Gene Expression Array Analysis

The health disparities that exist between minority women and CA women with TNBC are undoubtedly a result of a combination of factors: socio-economic, lifestyle, tumor characteristics, and inherent factors, such as genetic composition. Our group is focused on the genetic contributions to these disparities, to increase understanding of underlying biology, leading ultimately to individualized, ethnic-specific diagnostic and therapeutic approaches. In our pilot studies we have focused on gene expression profiling in a multi-ethnic BC cohort. Samples were obtained from archived FFPE blocks stored at the University of Miami, Department of Pathology and Kijabe Hospital in Kenya. Using 3-10μm scrolls from each block, tumor section and matching adjacent normal sections were macro-dissected, total RNA was isolated, cDNA prepared, and hybridized to an breast cancer enriched gene expression array (*Affymetrix Platform Breast Cancer DSA Research Tool*) in collaboration with Almac Diagnostics. A total of 60,856 gene/probes were analyzed and normalized using the Robust MultiArray Average (RMA) technique, which briefly, provides non-linear background correction on a per-chip basis, log transformed to the baseline median of all samples.

Table 3. Sample Ethnicity and Node Status that Passed QC

Ethnicity	Node status	Passed QC
Kenyan	Node 0	6
Kenyan	Mixed Node	15
African American	Node 0	10
African American	Mixed Node	7
Caucasian American	Node 0	13
Hispanic American	Node 0	12
Total		63

Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent

Gene expression analysis was conducted using *GeneSpring 12.1*® analytical software. After data QC analysis, it was determined that only those samples that passed quality control metrics to measure assay and hybridization performance would be used in further analysis - samples that passed are displayed in Table 3. Additionally, data showed that samples clustered well with respect to ethnicity (Figure 7). After quality control assessment of array data, filtered by expression, our multi-ethnic cohort for further analysis consisted of the following: 10-African American (AA), 12-Hispanic American (HA), 13-Caucasian American (CA) and 21-Kenyan. For the purposes of the comparisons presented here, only node 0, tumor vs. tumor comparisons are presented. A manuscript discussing all of these data sets is being prepared.

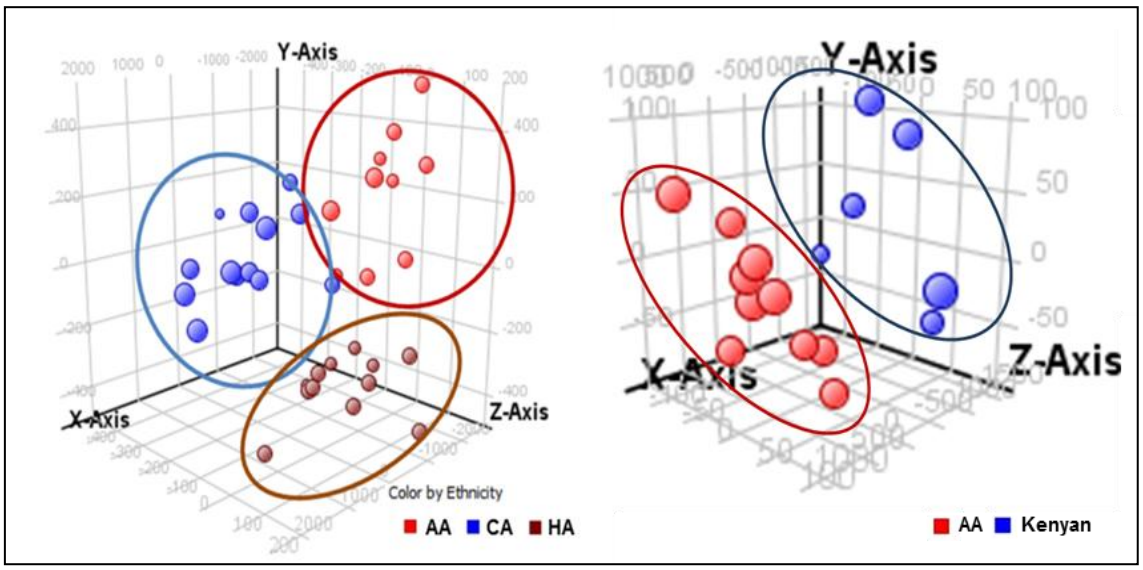


Figure 7. Principal Component Analysis: Ethnicity

Gene Clusters: TNBC Node 0 AA vs. CA

In our first level of analysis, we examined the normalized data for gene clusters between the AA and CA cohort. Unsupervised cluster analysis was performed using the hierarchical cluster algorithm, based on ethnicity and genes (p-value < 0.05, fold change > 2.5) and Pearsons uncentered similarity metric with centroid linkage rule. Based on gene expression profiling results, this revealed a majority of upregulated genes, found in the AA Tumor cohort compared to CA (Figure 8). Next, we identified differentially expressed genes between the two groups using a one-way ANOVA (p-value > 0.05) and fold change comparison (> 2.0). Significantly expressed genes, were determined after performing the Benjamin and Hochberg method for multiple-testing correction, which resulted in 128 statistically-significantly differentially expressed genes. **Interestingly, the list revealed significantly upregulated genes associated with the Wnt/β-catenin pathway in the AA cohort, as compared to the CA tumors (Table 4).**

This preliminary data supports the hypothesis that the Wnt/β-catenin pathway may contribute to a more aggressive TNBC phenotype in African American women. This pilot cohort will need to be further evaluated for molecular subtyping and TNBC classification. The upregulated genes in the AA cohort, associated with the Wnt/β-catenin pathway, will need to be validated *in silico* using much larger data sets. The genes set will also need to be evaluated *in vitro* and *in vivo*, and the functional relevance of

these genes in TNBC, particularly in an AA cohort, will need to be assessed. If confirmed in larger cohort series, these studies may have important implications for addressing BC ethnic disparities, as well as tailored approaches to prediction, prevention and treatment.

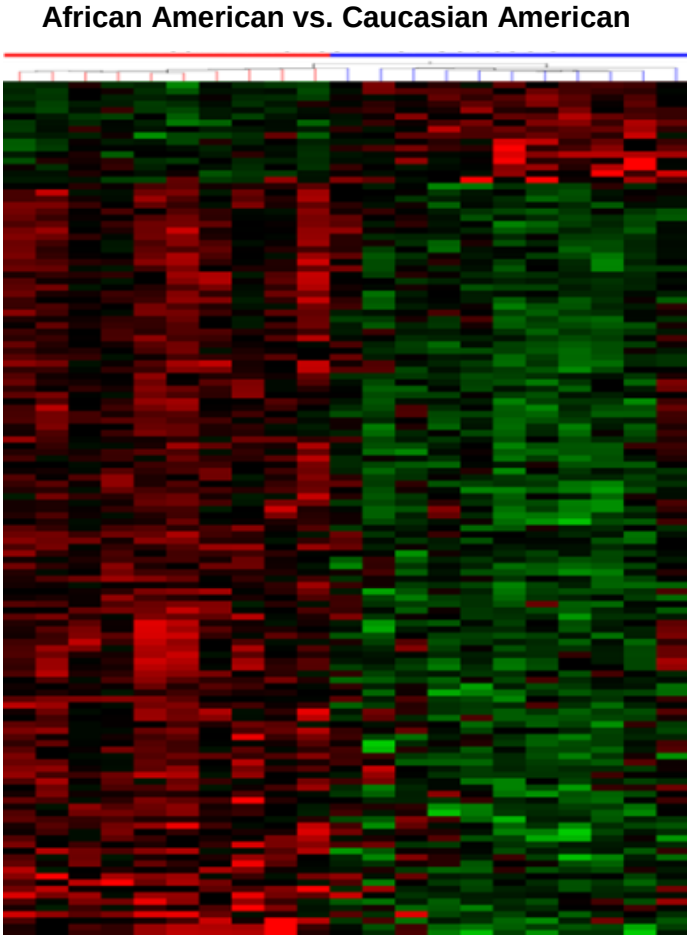


Figure 8. Hierarchical Clustering of AA vs. CA Triple Negative Breast Cancer Tumors. Clustered genes, p-value < 0.05 and fold change > 2.

Table 4. Genes associated with the Wnt/B-Catenin Pathway in AA vs. CA data set.

Gene Symbol	Gene Name	Gene Funtion	pValue	FC	Log FC
TCF4	Transcription factor 4	Binds to Wnt response elements to provide docking sites for β-catenin	0.001	3.34	1.74
CAV1	Caveolin 1	Wnt/β-catenin signaling/Epithelial mesenchymal transition–associated (EMT-associated)	0.006	3.29	1.72
FOXO3A	Forkhead box 3A	β-catenin binds directly to FOXO and enhances FOXO transcriptional activity	0.014	2.64	1.40
TNC	Tenascin-C	Down-regulation of the Wnt inhibitor Dickkopf	0.011	2.64	1.40

Gene Clusters: TNBC Node 0 Kenyan vs. AA

Next, unsupervised cluster analysis was performed using the hierarchical cluster algorithm, based on ethnicity and genes (p-value < 0.05, fold change > 2.5) and Pearsons uncentered similarity metric with centroid linkage rule. Based on gene expression profiling results, this revealed a pattern of differentially expressed genes in the Kenyan vs. AA Tumor cohort (Figure 9). We identified differentially expressed genes between the two groups using a one-way ANOVA *(*p-value > 0.05) and fold change comparison (> 2.0). Significantly expressed genes, were determined after performing the Benjamin and Hochberg method for multiple-testing correction, which resulted in 164 statistically-significantly differentially expressed genes. **Interestingly, the list revealed significantly deregulated genes associated with the Oncostatin M pathway in the Kenyan cohort, as compared to the AA tumors (Table 5).**

Table 5. Oncostatin M Signaling-associated genes in the Kenyan vs. African American TNBC Cohort.

Oncostatin M Signaling via MAPK	Fold Change
OSM Receptor	-5.67
JNK(MAPK8-10)	-4.14
STAT1	-17.02

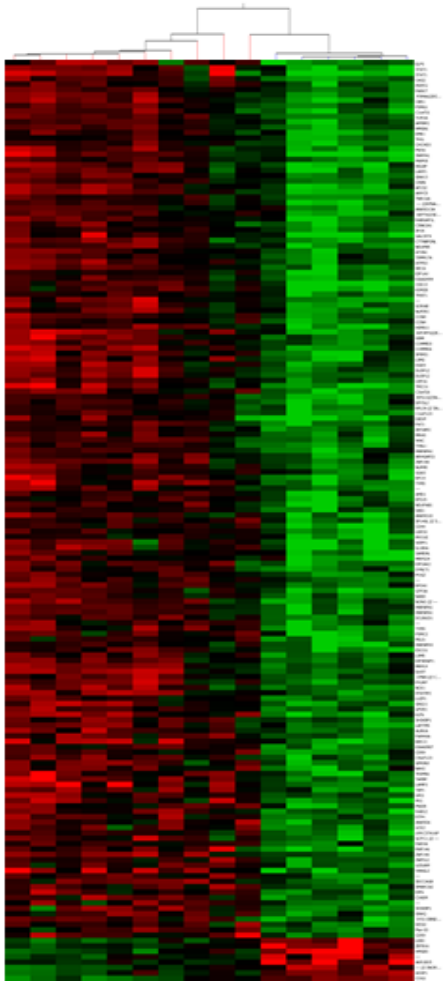
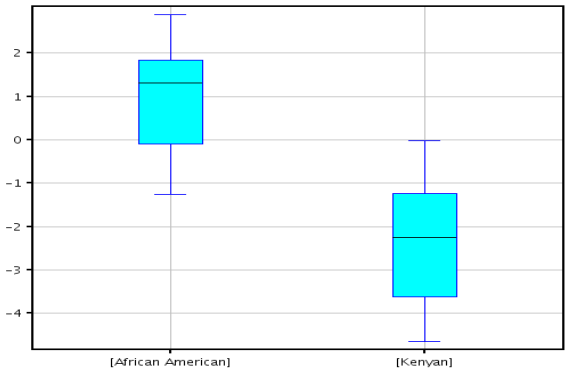


Figure 9. Hierarchical Clustering of AA vs. Kenyans TNBC.

Signal transducers and activators of transcription (STATs) are mediators of cytokine and growth factor receptor signaling, and STAT1 in particular, is the most deregulated gene in the Kenyan vs. African American cohort. As shown in Figure 10, STAT1 is significantly downregulated in the Kenyan cohort, compared to the African American. Dysregulation of STATs has been implicated in cancer, and STAT1 has been shown to function as a tumor suppressor, and over-expression correlates with an overall better prognosis in breast cancer. Additionally, STAT1 loss has been shown to cause mammary cancer initiation and growth in mice (7). These finding will require further investigation; however, STAT1 could play a role in the health disparities that exist in TNBC among the Kenyan, AA and CA populations.



Displayed are the overall fold gene expression differences for STAT-1 measured between all AA and Kenyan samples, incorporating all samples passed QC and all arrays.

Figure 10. STAT1 expression: African American vs. Kenyan

3) KEY RESEARCH ACCOMPLISHMENTS

- Task 1: Extraction and preparation of DNA and RNA from FFPE tumor samples from North American African and Kenyan African cohorts.

- Task 2: Aim I, Analyze and compare genome-wide transcript expression in BC samples of AA ancestry vs. native African (Kijabe) samples.

4) REPORTABLE OUTCOMES

- **Continued Identification of Ethnic Specific Differences in Breast Tissue – on the Road to Biomarker Discovery in Breast Cancer.** Lisa L. Baumbach-Reardon, Mary Ellen Ahearn, Carmen Gomez, Aldo Mejias, Merce Jorda, Tom Halsey, Jim Yan, Kevin Ellison, Karl Mulligan, Mark Pegram. Univ. of Miami Medical School, Miami, FL, Almac Diagnostics, Durham, NC. *AACR Special Conference*, *The Future of Molecular Epidemiology: New Tools, Biomarkers, and Opportunities*, June 6 – June 9, 2010, Miami, Florida.

- **Genomic/genetic differences in breast cancer across ethnicities.** Lisa L. Baumbach-Reardon. University of Miami **Platform Presentation/ Invited Speaker; AACR** *Conference on The Science of Cancer Health Disparities in Racial/Ethnic Minorities and the Medically Underserved,* September 30-October 3, 2010.

- **Continued Identification of Ethnic Specific Differences in Breast Cancer and Normal Breast Tissue.** L. Baumbach, M. E. Ahearn, C. Gomez, A. Mejias, M. Jorda, T. Halsey, J. Yan, K. Ellison, K. Mulligan, R. Kittles, A. Ashworth, M. Pegram Univ Miami School of Medicine, Miami, FL; Almac Diagnostics, Durham, NC; University of Illinois at Chicago, Chicago, IL; Breakthrough Cancer Research Center, London, UK. American *Society of Human Genetics Annual Meeting,* November 2010.

- **Gene Expression Profiling of Formalin-Fixed, Paraffin-Embedded (FFPE) Breast Cancer Samples and Analysis of Intrinsic Subtypes** Baumbach LL, Gomez C, Yan J, Halsey T, Ahearn ME, Jorda M, Kennedy R, ODonnel J, McDyer F, Deharo S, Pegram M. University of

Miami Medical School, FL; Almac Diagnostics, Durham, NC**.** *San Antonio Breast Cancer Symposium,* December 2010.

- **Identification of ethnic specific differences in breast cancer and normal breast tissue** Lisa L. Baumbach, Carmen Gomez, Jim Yan, Tom Halsey, Mary Ellen Ahearn, Merce Jorda, Mark Pegram. University of Miami School of Medicine, Miami, FL; Almac Diagnostics, Durham, NC American *Association for Cancer Research Annual Meeting (highlighted for media attention)*, Orlando, FL (2011).

- **Investigation of Transcriptome Differences in Breast Cancer Tissues from African-American and East African Patients with Triple Negative Breast Cancer**. Lisa Baumbach, Translational Genomics Research Institute, Phoenix, AZ **Platform Presentation/ Invited Speaker;** AACR *Conference on The Science of Cancer Health Disparities in Racial/Ethnic Minorities and the Medically Underserved,* October 27-30, 2012.

- **Comparison of transcriptional signatures in US African American and Kenyan TNBC samples identifies differential expression in key oncogenic pathways**. Baumbach-Reardon LL, Getz JE, Ahearn ME, Gomez C, Bird P, Carpten J, Pegram M. *American Association for Cancer Research Annual Meeting,* Washington, D.C. (2013).

5) CONCLUSION

RNA and DNA extracted from these samples are usually degraded, contaminated and of low quality in general. Despite the large banks of FFPE samples available for retrospective studies that include follow-up analysis of patient outcome, most of these studies currently focus on frozen samples because of the limited options available for paraffin samples. Additionally, FFPE processing holds advantages for tissue storage during prospective studies, in which many biopsies are collected but only a fraction of them are applied to downstream assays with selection based on clinical outcome. Because of the difficulty and time required to obtain fresh frozen tumor samples from the triple negative breast cancer patients with matched clinical criteria and curation, this study explored the possibility to profile both gene expression and genotype from FFPE tumor tissues. This study attempts to test and establish the feasibility and outline guidelines for selection of technology platforms and QC criteria for FFPE samples. FFPE RNA and DNA that are applied to the Almac Diagnostic Breast Cancer DSA arrays may still vary in quality and therefore require careful and rigorous QC to select samples that meet the quality standard including chip CQ and sample integrity check at profiling level. In our CNV data, the QC performance of FFPE sample is not comparable to fresh frozen samples. Consequently, even with careful specimen processing (including multiple attempts at DNA extraction in three different laboratories), QC and data analysis, information such as LOH, and copy number assessment could not reliably be obtained from the Kijabe Kenyan African cohort, presumably due to *pitfalls in pre-analytical sample handling in the field*, resulting in significant DNA degradation. These results underscore continuing challenges for the application of FFPE samples to the same genome-wide platforms already available for high-quality DNA samples. Nevertheless, our preliminary transcript array data supports the hypothesis that the Wnt/β-catenin pathway may contribute to a more aggressive TNBC phenotype in African American women. Moreover, we find that STAT1 is significantly downregulated in the Kenyan cohort, compared to the African American samples. While these finding will require further investigation, these data support our hypothesis that comparison of transcriptional signatures in US African American and Kenyan TNBC samples identifies differential expression in key oncogenic pathways which play an important role in the health disparities that exist in TNBC among the Kenyan, AA and CA populations.

6) REFERENCES

1. American Cancer Society (2007) Cancer Facts and Figures for African-Americans.
2. Chlebowski RT et al (2005) Ethnicity and breast cancer: factors influencing differences in incidence and outcome. *JNCI* 97:6, 439-48.
3. Albain KS et al (2003) Outcomes of African Americans in SWOG BC therapy trials.*Br Can Res Treat* 77:A21.
4 Shen Y et al (2007) Are there racial differences in breast cancer treatments and clinical outcomes for women treated at M. D. Anderson Cancer Center? *Breast Cancer Res Treat* 102:347–356.
5. Bird PA, Hill AG, Houssami N. (2008) Poor hormone receptor expression in East African Breast Cancer: Evidence of a Biologically Different Disease? *Annals of Surgical Oncology*
6. Arriola E, Marchio C, Tan DS, et al. (2008) Genomic analysis of the HER2/TOP2A amplicon in breast cancer and breast cancer cell lines**.** *Lab Invest* 88: 491-503.
7. Schneckenleithner C, Bago-Horvath Z, Dolznig H, Neugebauer N, Kollmann K, Kolbe T, Decker T, Kerjaschki D, Wagner K, Müller M, Stoiber D and Sexl V. 2011. Putting the brakes on mammary tumorigenesis: Loss of STAT1 predisposes to intraepithelial neoplasias. Oncotarget 2(12), pp. 1043-54.

7) APPENDICES

Appendix 1. Research Personnel. Pursuant to the USAMRMC technical reporting requirements, below are all personnel involved in the research. The budget was modified in 2012 to accommodate the increased effort of Dr. Mark Pegram and Rebecca Olson for the final calendar year of the project.

Table 6. List of Personnel Receiving Pay from the Research Effort

Name	Role
Mark Pegram	principal investigator
Lisa Baumbach	principle investigator
Jennifer Clarke	biostatistician
Catherine R. Connor	consultant
Peter Bird	co-investigator, subcontract
Alan Ashworth	subcontract
Jorge Reis-Filho	subcontractor
Rebecca Olson	research assistant